# 基于图像处理和复杂网络分类方法的新冠肺炎诊断软件使用说明

## 功能概述

软件用途：本软件主要是通过对人的胸腔X光图像进行分析，然后运用分形维数提取图像的特征，最后利用复杂网络的方法对人的胸腔X光图像进行分类，最后达到诊断感染新型冠状肺炎患者的目的。

技术特点：感染了新型冠病毒的人的胸部X射线图像呈现几何形变，例如丝状纹理的形成，该技术针对此问题，提出我们在不同阈值的二值图像上计算图像的分形维数 [1]，来表征COVID-19图像的复杂度，然后运用基于数据的联系结构分类技术对图像进行分类。

开发语言及环境配置

开发语言是python

matplotlib 3.5.1

networkx 2.6.3

numpy 1.21.5+mkl

opencv-python 4.5.5

pandas 1.3.5

python 3.8.11

scikit-learn 1.0.2

### 理论概述

传统的分类技术，例如k最近邻（kNN），贝叶斯决策理论，神经网络，决策树等, 在本质上，所有这些技术都根据输入数据的物理特征（例如，距离或相似性）训练未分类的数据项并对其进行分类。此外，科员人员已经设计出了其他一些利用类拓扑特性进行预测的监督数据分类技术，例如支持向量机（SVM）和线性判别函数。以支持向量机为例，在线性情况下，支持向量机用凸包找到最佳的分离超平面。在非线性情况下，数据被映射到更高维的空间，并且通过选择适当的内核，在映射空间中生成线性凸包近似于真实的非线性类拓扑。这些经常使用的内核包括径向基函数（RBF），sigmoid函数，多项式和线性核函数。这些使用物理特征或类拓扑而不是类模式形成的技术对数据进行分类的方法成为传统的分类技术。

通常，数据项不是属性空间中的孤立点，而是倾向于形成某些模式。例如，在图2.1中，如果仅考虑物理特征（例如数据实例之间的距离），则由“三角形”（黑色）表示的测试实例最有可能被归类为“正方形”（蓝色）类的数据项；另一方面，如果考虑到数据项之间的语义含义，将直观地看到“房屋”轮廓的清晰图案，因此将“三角形”项目分类为“圆形”（红色）类的数据项。人（动物）的大脑脑执行低阶和高阶学习，并且具有根据输入数据的语义识别模式的能力，但是，一般来说，这种任务仍然难以通过计算机进行评估。在这里，不仅考虑物理属性和类拓扑而且还考虑模式形成的监督数据分类被称为基于数据的联系结构的分类技术[2]。这项技术初始是Silva和Zhao在2018年的一项工作中提出来的。换句话说，传统的一些数据分类比如说KNN，它们根据训练数据的物理特征来进行训练模型和分对新数据进行类，从而忽略了数据项之间的许多其他固有和语义关系，通常会在数据空间中生成复杂形式的类；另一方面，已知人（动物）大脑可以根据输入数据的语义来识别模式。在这种情况下，基于网络的技术可以从截然不同的观点做出贡献。

在本软件中，我们对Silva和Zhao 技术进行了改进。具体讲，我们用可通信性这样一个网络刻度来描述网络特征；而在Silva和Zhao的技术中，是用三个刻度来描述网络特征。由于我们只用一个网络刻度，就避免了选取网络刻度权重这样一个复杂的优化问题，同时也得到高精度的分类结果。

图表, 散点图

描述已自动生成

图2.1 监督数据分类任务的简单示例

基于数据的联系结构分类技术在接下来的工作中进行了扩展。在该方案中，底层分类可以通过任何传统的分类技术来实现，而高层分类可以通过任何传统的分类技术来实现探索了由输入数据建立的网络的复杂拓扑属性。而该软件中就是使用的改进的基于数据的联系结构分类技术对新冠感染者的胸腔X光片进行分类。

### 3．数据集说明

程序中的所有路径中，NORMAL文件夹是正常人的胸腔X光片原始图像数据集，COVID-19文件夹是感染了新型冠状病毒的胸腔X光片原始图像数据集，最后构建的数据集总量是300张图像，每个类别150张；NORMAL1文件夹是正常人的胸腔X光片分析样本图像数据集，COVID-191文件夹是感染了新型冠状病毒的胸腔X光片分析样本图像数据集。

### 4．使用说明

程序分为3个文件夹，这三个文件夹是COVID-19-c,fractral dimensionc和model。 fractral dimension下的dimensions是画出fractral dimensions折线图；model文件夹下的build\_network是构建分类网络，GetCOVID\_19Data1是提取图像特征，main是主程序;COVID\_19-c是图像数据。

#### 4.1 图像特征提取及分类

进入fractral dimension文件夹，运行dimension lines.py文件，会生成样本数据集中的8张图像的不同阈值对应的二值图像，图4.1.1为4张正常人的胸腔X光片，图4.1.2为4张CONVID-19的原图及不同阈值下的二值图像；图4.1.3是对应的分形维数的折线图，其中（a）-（d）四条折线是正常人的胸腔X光片对应的分形维数，（e）-（h）四条折线是感染了新型冠状病毒的人的胸腔X光片对应的分形维数。

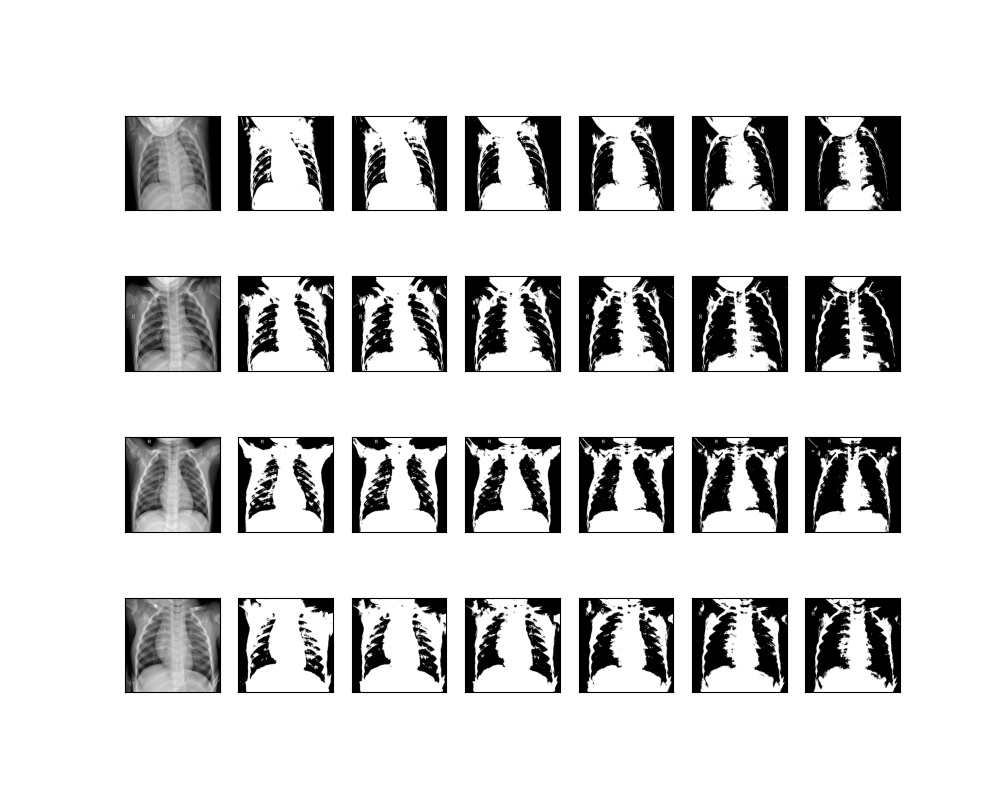


图4.1.1 正常人的胸腔X光片图像及其对应的二值图像

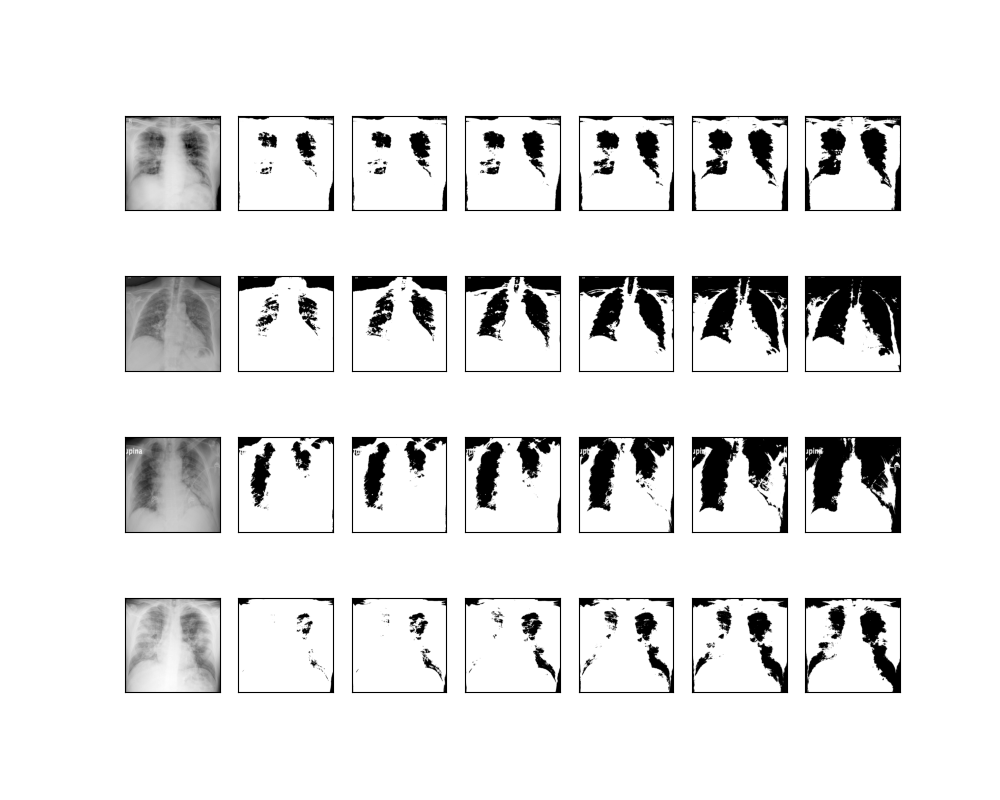


图4.1.2 感染新冠的人的胸腔X光片图像及其对应的二值图像

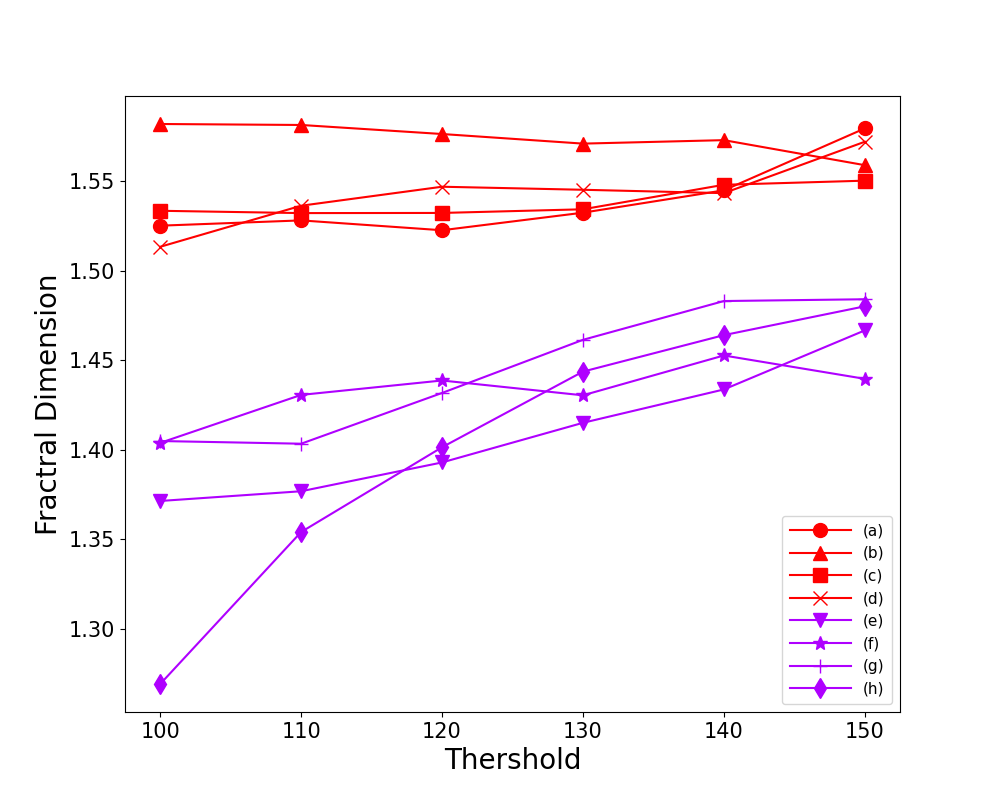


图4.1.3 分形维数折线图

进入model文件夹，运行main.py文件，会生成初始化网络节点图和分类节点图，如图4.1.4和图4.1.5所示。网络节点图中红色节点表示正常类别，蓝色节点表示新冠类别。图4.1.5中，黑色节点表示新插入的数据节点。

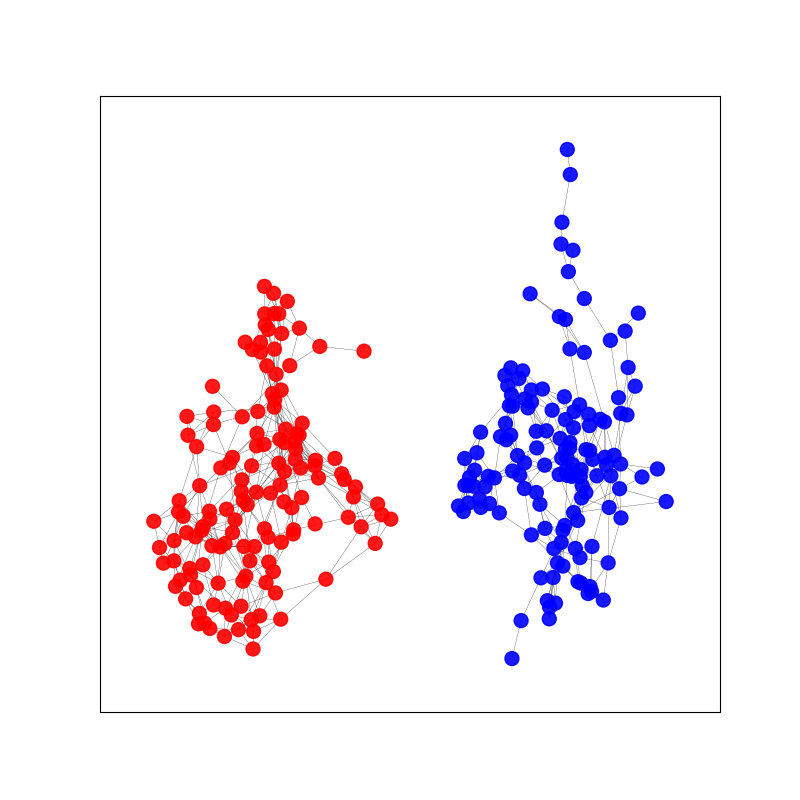


图 4.1.4初始化网络节点图

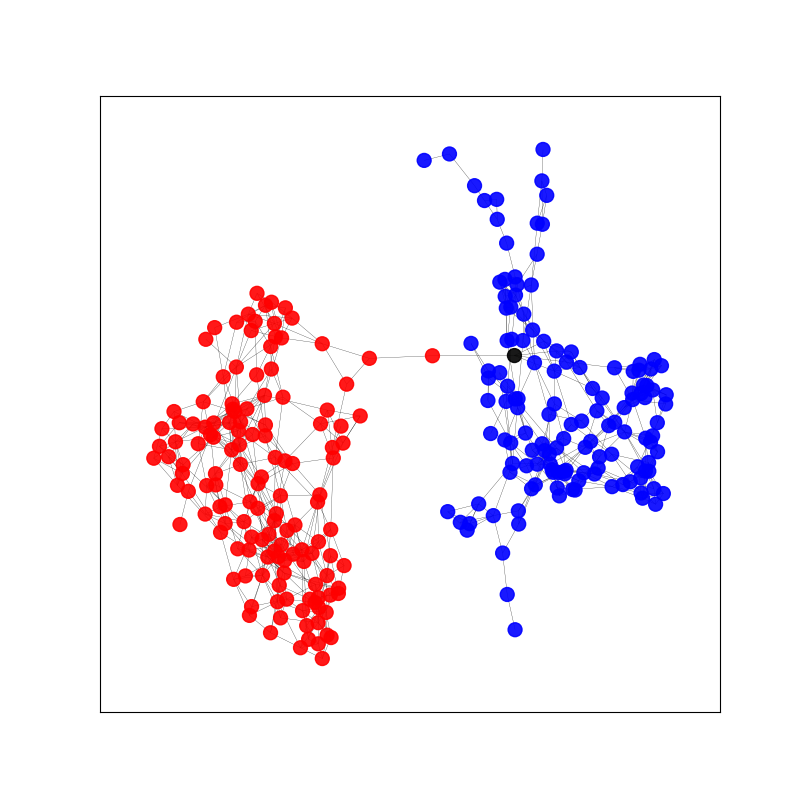


图4.1.5 分类节点图

在分类阶段如图4.1.5，将测试数据样本图像（黑色节点）插入到训练集构建的初始化网络中，网络构建完成后，计算每一个网络的度量，即，计算。在该技术方法中，使用平均通性度量作为，它不仅说明了连接两个节点的最短路径，而且说明了贡献较低的较长路径。

从节点到网络所有其他节点的可通信性

 （4.1.1）

其中是和之间的最短路径的长度，是和之间的最短路径的数量，是大小为的连接和的路径的数量。这种选择背后的原因是，最短路径会受到网络结构变化的显著影响 [3]。

将未标记的数据样本一一分类。首先，将计算*X*items中的所有数据插入每一个网络后形成的网络度量的平均值，这在第一步中已经存储在了二维数组中；然后，将*X*test中的一个新数据样本插入到目前已构建的两个网络中。然后，计算插入后每个网络的相同度量。 最后，通过公式4.1.3得到了新样本插入每个类的影响：

 （4.1.2）

最终这个新的样本数据分类到类中，其中：

 （4.1.3）

换句话说，新样本符合网络形成的模式，并且不会对网络产生较大的干扰。请注意，新样本甚至可以与类的元素相距甚远。

**References**

[1] Benoit Mandelbrot, How Long is the Coast of Britain? Statistical Self-Similarity

and Fractional Dimension, Science, 156, 636-638 (1967)

[2] T. C. Silva and LIANG ZHAO, “Network-Based High-Level Data Classification”. IEEE Transactions on Neural Networks and Learning Systems, vol. 23, pp. 954-970, 2012.

[3] Estrada E. and Hatano N.: Communicability in complex networks, Phys. Rev. E,

77, 036111 (2008). DOI: ”https://doi.org/10.1103/PhysRevE.77.036111”